Variability of Crustacean Zooplankton Data Generated by the Great Lakes National Program Office's Annual Water Quality Survey

Richard P. Barbiero

DynCorp Science and Engineering Group
1359 West Elmdale Avenue
Suite #2
Chicago, Illinois 60660

Prepared for:
United States Environmental Protection Agency
Great Lakes National Program Office
77 West Jackson Boulevard
Chicago, Illinois 60604

Louis Blume, Project Officer

September 2003

Acknowledgements

This report was prepared under the direction of Louis Blume, Project Officer, Great Lakes National Program Office (EPA Contract No. 68-C-01-091). Assistance with ANOVA calculations was provided by Ken Miller, DynCorp Science and Engineering Group.

Summary

A Data Quality Objective (DQO) has been developed by the Great Lakes National Program Office (GLNPO) to ensure that data collected from their Water Quality Surveys are of suitable quality to provide decision makers with sufficient certainty to make educated ecological management decisions. The current GLNPO DQO states that data quality should be sufficient for there to be an 80% chance of detecting a 20% change, at the 90% confidence level, between current and historical measurements of a variable made in a particular lake during a particular season.

This report determines the extent to which zooplankton data comply with the GLNPO DQO, and assesses the relative contribution of different sources of variability to the overall uncertainty of zooplankton data. The most important findings are summarized below:

- Data quality of zooplankton data falls far short of the current DQO. In only 3 of 184 cases examined was the DQO criterion met.
- Minimum detectable differences for the major taxonomic groups and the most common species were largely between 40 and 190%.
- Estimates of cladoceran densities were most variable; estimates of calanoid copepod densities were least variable.
- It is unclear if the current data quality is sufficient to detect ecologically important trends. A recent study show that in at least some cases it is (Barbiero and Tuchman, in press).
- Relatively little variability is due to analyst error in counting/identification.
- About 25% of variability is introduced during the field sampling and/or laboratory subsampling stages.
- The majority of uncertainty in zooplankton data is due to station-to-station (within basin) variability. Reducing this source of variability would entail increasing the number of sampling stations.
- The most practical way to reduce variability is to ensure proper functioning/reading of the flow meter.
- Since the variability introduced into the analysis by subsampling in the laboratory is unknown, a study quantifying this source of uncertainty could point to further means of reducing variability.
- An appropriate QC criterion for relative species composition of duplicate laboratory analyses, using the PSc index, is 0.92.
- An appropriate RPD QC criterion for total organism counts in duplicate laboratory analyses is 4%.

1 Introduction

1.1 GLNPO water quality survey

The Great Lakes National Program Office (GLNPO) of the U.S. EPA has been involved in regular surveillance monitoring of the open waters of the Laurentian Great Lakes since 1983. This surveillance monitoring is meant to satisfy the provisions of the Great Lakes Water Quality Agreement (International Joint Commission 1978), which calls for periodic monitoring of the lakes to evaluate the effectiveness of pollution control/reduction strategies in the Great Lakes, recognize emerging problems, and identify the need for new or revised strategies and further research. According to GLNPO (2003), the water quality surveys have been specifically designed to:

- Focus on key physical, chemical, and biological indicators of lake health
- Evaluate the health of each lake under different conditions (stratified and unstratified)
- Allow for real-time detection of significant changes in water quality, as indicated by significant changes in one or more parameters
- Provide data that can be compared from year to year
- Provide data to support decisions regarding the need for further study or new pollution control strategies

In order to ensure that data collected from GLNPO's water quality surveys fulfill these requirements, a data quality objective (DQO) has been developed to be applied to all water quality survey data. Management of data quality is an important aspect of the larger mission of the water quality surveys, and requires an understanding both of the overall magnitude of variability, and of the relative contributions of individual components of sample collection and analysis to total variability. More fundamentally, it is also necessary that the DQO be sufficiently explicit to enable its unambiguous application to water quality survey data, and that it be appropriate to the type of data collected by the water quality survey.

Recognition of the importance of open water planktonic communities in the overall assessment of ecosystem health led to the inclusion of sampling for zooplankton communities at the inception of the monitoring program. However, data generated from the sampling of biological communities poses special challenges for the application of the DQO and for assessments of variability. DQOs are typically developed in relation to chemical variables, which are characteristically univariate, unlike biological community data, which are multivariate. It is important, therefore, to assess both the extent to which the DQO is applicable to biological data, and whether or not that data satisfies the DQO.

1.2 Objectives of study

The overall purpose of the present study was to provide an assessment of the variability of data generated by GLNPO's zooplankton monitoring program. The specific goals of the study were several fold:

1. To determine the minimum detectable differences under the current sampling regime;
2. To determine if the current level of effort satisfies the GLNPO DQO;
3. To determine the relative contribution to overall variability of different stages of sample collection and analysis;
4. To determine appropriate analysis criteria for duplicate laboratory (QC) analyses.

In addition, the applicability of the current DQO to zooplankton data is discussed.

2 GLNPO's Data Quality Objectives

2.1 Current ambiguities

The assessment of lake health using data generated from the water quality survey requires that sufficient data quality be obtained to permit detection of 'significant' changes in the variables under consideration. For the purposes of the water quality surveys, GLNPO has defined a significant change as a 20% difference between current and 'historical' measurements, made for a particular variable in a particular lake during a particular season. The DQO for GLNPO's water quality survey is stated as the ability to "collect measurements that will yield an 80% chance of detecting a change of 20% or more within a particular lake and season, at the 90% confidence level" (p. 15; GLNPO 2003). This formulation of the DQO, however, contains several ambiguities, particularly as it relates to multivariate data such as that generated from zooplankton analyses. First, as currently stated the DQO does not indicate what the detection target of a 20% change is in relation to. Elsewhere in the same document, both a comparison to 'historical' values (p. 15; GLNPO 2003) and comparisons between two years (p. 27; GLNPO 2003) are referred to. As pointed out elsewhere (Barbiero, 2003), the detection of a change between a given season's data and 'historical' values can be variously interpreted to mean a change in relation to the previous year's data, a change in relation to a pooling of all previous years' data, or a change in relation to any previous year's data. An additional possible interpretation of the DQO would be to permit the detection of a *trend* in historical data, although this would not seem to be completely consistent with its current formulation.

The DQO also appears to be at variance with the basic statistical design of the water quality surveys, in that the target change is defined in the DQO on a lake-wide basis, while the statistical design of the survey is based on replication at the level of two or three homogeneous basins within each lake (p. 27, GLNPO 2003). This can be accommodated for by employing a stratified statistical design in assessing changes in variables, i.e., by first computing the values of each variable on a basin-wide basis, and then combining those estimates in proportion to how much of the lake each basin accounts for to arrive at a lake-wide estimate. Under this scenario, variance would also have to be calculated proportionately. Interpreting the DQO in this way, however, assumes that changes can only take place on a lake-wide basis. In a case where the timing and/or magnitude of change differed from basin to basin, as for instance might be expected in Lake Erie where differences in morphometry result in vast differences in the chemical and biological characteristics of the three basins, limiting the detection of changes to a lake-wide basis could obscure changes taking place only within a given basin.

While it is not within the scope of this report to clarify the ambiguities of the current GLNPO DQO, in order to apply it to the zooplankton data, some assumptions had to be made concerning its interpretation. For the purposes of this report, the DQO was assumed to denote the requirement of an 80% chance of detecting of a 20% change *between two years within a given basin* for a particular season at the 90% confidence level.

2.2 Application to multivariate data

Resolving the ambiguities in the current formulation of the DQO is theoretically possible. More fundamental difficulties exist, however, in the application of the DQO to data generated by the zooplankton sampling program. As with all data generated by the biological monitoring program, zooplankton data are multivariate. Each sample, rather than producing a single value associated with a single variate, will produce values associated with a varying number of variates. Variates here correspond to the different species identified in each sample, and values associated with those variates correspond both to the densities of those species, and to their biomass. The variates produced by a sample will not necessarily be consistent within groups of replicates, nor will they even necessarily be the same between replicate analyses of the same sample. Theoretically, then, the DQO could apply to each individual variate (i.e., species) identified within a sample. A given sample could therefore be called upon to satisfy as many DQOs as there are species within that sample, which in the case of zooplankton could be expected to vary between several and several dozen. In addition, it might be of interest to assess changes in broader taxonomic categories of organisms, for example to assess changes at the taxonomic level of order or suborder (e.g., cladocerans, calanoid copepods, etc.), or to assess changes in various functional groups (e. g., grazers, predators, etc.), or indeed to track changes in total zooplankton density or biomass.

One problem, therefore, arising from the multivariate nature of zooplankton data is deciding upon the variate(s) of interest. It is likely that changes in the populations of some species, or certain groupings of species, are of little inherent ecological interest, and therefore do not need to be subject to the DQO. Also, the statistical difficulties associated with estimating the abundances of species that typically occur in very small numbers might preclude their ability to conform to the DQO.

A more fundamental problem exists, however, if community-level attributes of the zooplankton data are of interest. Examination of overall community structure often reveals changes that are not apparent from examination of individual species (Yan *et al.*, 1996), and could provide a more relevant measure of ecosystem health. In this instance, defining an appropriate metric, and quantifying the variability associated with that metric, becomes highly problematic. Changes in community structure are typically quantified using multivariate techniques, but metrics derived from such techniques are often not easily convertible into a single number, nor are there universally accepted methods of quantifying the variance of such metrics, and they thus would not be easily amenable to assessment in terms of the current DQO. There are currently no guidelines in place to enable the application of the GLNPO DQO to multivariate community level data.

3 Zooplankton Program

3.1 Overview

GLNPO's regular surveillance monitoring of the open waters of the Laurentian Great Lakes began in 1983. Initially, only the open waters of Lakes Michigan, Huron and Erie were included in GLNPO's monitoring program. In 1986, monitoring of Lake Ontario was added, and in 1992, Lake Superior was included.

Sampling protocols have undergone some changes since the beginning of the program. In 1983 and 1984, two vertical zooplankton tows were taken at each site with a 63-µm mesh net: one from 2 m above the bottom to the surface, and a second from 20 m to the surface (Makarewicz, 1987; Makarewicz, 1988). In 1985, the deeper tow was apparently discontinued (Makarewicz and Bertram, 1991), leaving just the 20-m tow. Concerns about the representativeness of samples collected from just the upper 20 m of the water column led to a further change in the zooplankton sampling protocol. Starting in the summer of 1997, a second tow was added to the sampling regime. This tow was taken from a depth of 100 m, or 2 m from the bottom, whichever was shallower. Unlike previous deep tows, the 100-m tows were taken using a net with a larger mesh size (153-µm) to prevent clogging and to reduce the pressure wave created by the net during sampling. Also, the time of day at which the tows were taken was recorded from 1996 on, something which had not been done earlier.

There are two main consequences of taking zooplankton tows from relatively shallow depths. In species that undergo diurnal vertical migration, 20-m tows taken during the day, when such species are typically below the epilimnion, can result in an underestimation of abundances. This would lead both to unrepresentative samples, and also to an increase in both inter- and intra-annual variability. If replicate sites are sampled at different times of day during a cruise, as is often the case, intra-annual variability would increase, while if sites are visited at different times of day from year to year, as is also likely, this would result in an increase in apparent inter-annual variability. Secondly, populations of deeper-living zooplankton that rarely migrate above 20 m would be consistently underestimated in 20-m tows, whether taken during the day or at night. Because of the problems inherent in the interpretation of shallow, 63-µm mesh tows, emphasis in this report will be on the deeper, 153-µm mesh tows.

3.2 Field methods

Currently, two sampling tows are performed at each station. The first tow is 20 meters below water surface using a 63-µm mesh net. The second tow is a 'full' water column tow, to 2 meters above the bottom of the lake or 100 m, whichever is less, using a 153-µm mesh net. If the station depth is less than 20 m, both tows are taken from one meter above the bottom. Tows are taken with a 0.5-m diameter conical net (D:L=1:3) equipped with a flowmeter. Once on station, the biology technician resets the flowmeter dials to zero, and has the winch operator lower the net so the rim of the net is at the surface of water. The net is then lowered to the appropriate depth as indicated by a winch meter on deck, and raised it at a constant speed (at or close to 0.5 meter/second) until the rim of the net is approximately eye-level. Upon retrieval the flowmeter meters are read and the net is rinsed with a hose from the outside to wash all of the organisms off of the net cloth inside and into the sample bucket. The sample is concentrated into the sample bucket, which is then detached from the net and its contents rinsed and poured three times into a pre-labeled 500-mL sample bottle. The organisms are then narcotized with soda water and preserved with sucrose formalin solution. Triplicate tows of each depth are taken at the master stations.

3.3 Laboratory methods

Microcrustacea are examined in four stratified aliquots under a stereoscopic microscope. The sample is subsampled using a Folsom plankton splitter, with half of each split set aside, and the other half returned to the splitter to be split again. Successive splits are made until the last 2 subsamples contain between 200 and 400 microcrustaceans each (not including nauplii). In total, four subsamples are examined and enumerated. Each is removed, in turn, with a condensing tube and placed in a circular counting chamber. All microcrustaceans within each subsample are identified and enumerated under a stereozoom microscope. The four subsamples are: the final two, most dilute subsamples which contain 200-400 organisms, in which all microcrustaceans are examined and enumerated; a third subsample equal in fraction to the sum of the first two subsamples, which is examined for subdominant taxa (taxa enumerated less than 40 times in the first two subsamples combined); and a fourth subsample equal in fraction to the sum of the first three from which rare taxa are enumerated. In general, ten percent of all samples analyzed are analyzed in duplicate by a second analyst. If a given lake/cruise has less than 10 samples, at least one sample from that data set is also analyzed in duplicate. Duplicate analyses are performed after subsamples are placed into the counting chamber, and thus quantify variation associated with enumeration and identification, but not with subsampling.

4 Sources of Variability

4.1 Levels of replication

The statistical design of the zooplankton program follows that of the broader water quality monitoring program, with each lake divided into statistically homogeneous basins (Fig. 1; Table 1). Within each basin stations function as replicates, and provide an indication of large-scale spatial heterogeneity. Each basin contains a master station, usually located at the deepest point in the basin, at which triplicate zooplankton tows are taken. These tows function as field replicates and are meant to quantify the variability 'within' each station associated with sample collection, including variability associated with lowering and raising the net, the angle and actual (as opposed to nominal) depth of the tow, the functioning/reading of the flow meter, and the washing of the net bucket contents into the collection bottle. These field replicates also capture the variability due to smaller scale zooplankton patchiness.

In the laboratory each sample is subsampled, and subsamples from successive dilutions are counted to ensure accurate estimation of rarer species. There is no replication at this stage, so there is no way to estimate the amount of error introduced into the analysis by subsampling. The entire contents of each of four sub-samples are placed successively into the microscope chamber and identified and enumerated by the analyst. A second analyst pro-

Table 1. Assignment of GLNPO water quality survey stations to homogeneous basins with the five Laurentian Great Lakes.

Lake	Basin	Stations
Michigan	southern lake	MI 11, MI 17, MI 18, MI 19 MI 23
	central lake	MI 27, MI 32, MI 34
	northern lake	MI 40, MI 41, MI 47
Huron	northern lake	HU 45, HU 48, HU 53, HU 54, HU 61
	central lake	HU 32, HU 37, HU 38
	southern lake	HU 06, HU 09, HU 12, HU 15, HU 27, HU 93
Erie	western lake	ER 58, ER 59, ER 60, ER 61, ER 91, ER 92
	central lake	ER 30, ER 31, ER 32, ER 36, ER 37, ER 38, ER 42, ER 43, ER 73, ER 78
	eastern lake	ER 09, ER 10, ER 15, ER 63
Ontario	western lake	ON 12, ON 25, ON 33, ON 41
	eastern lake	ON 49, ON 55, ON 60, ON 63
Superior	western lake	SU 15, SU 16, SU 17, SU 18, SU 19
	central lake	SU 06, SU 07, SU 08, SU 09, SU 10, SU 11, SU 12, SU 13, SU 14
	eastern lake	SU 01, SU 02, SU 03, SU 04, SU 05

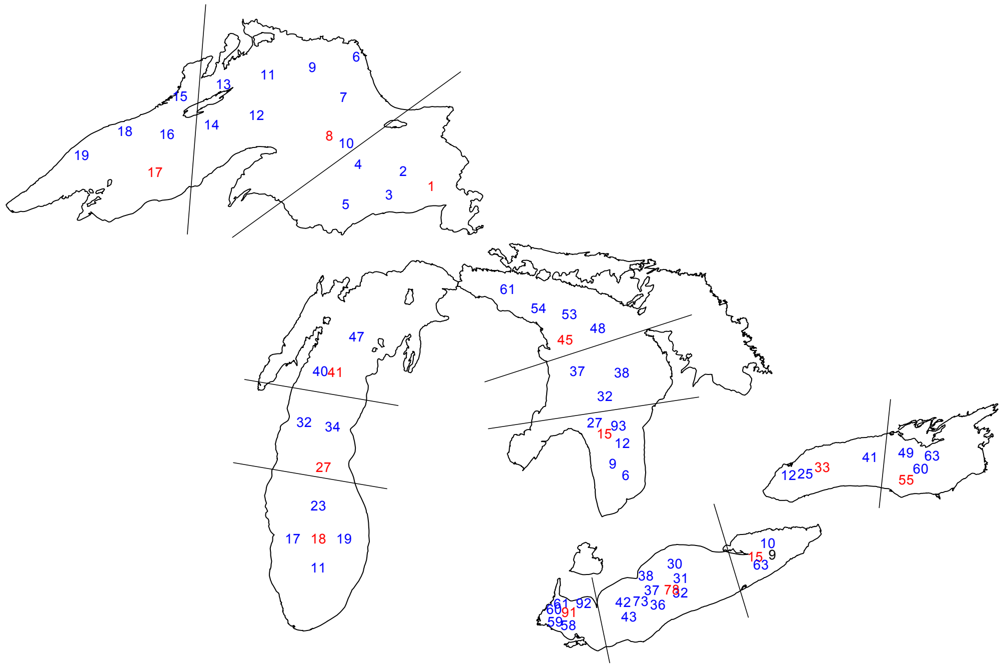

Fig. 1. Locations of GLNPO's water quality survey (WQS) sampling stations within homogeneous basins, as defined by 2003 quality assurance program plan. Master stations indicated in red.

vides duplicate counts and identifications of 10% of the samples. Duplicate analyses capture variability associated with species identifications and with the counting of animals within the chambers. These duplicate analyses are conducted after the subsamples are placed into the counting chambers, so as noted, no estimate of subsampling variability is possible. A summary of the main sources of variation is given in Table 2, along with the measures currently in place to estimate their magnitude.

4.2 Compliance with DQO

Assessing the degree to which the current sampling effort satisfies the DQO required that some assumptions be made in order to resolve the ambiguities in the DQO pointed out in Section **2.1**. As stated earlier, it was assumed that the DQO required data of adequate quality to permit an 80% chance of detecting a 20% change in a given variable *between two years* within a given *basin* and season with 90% confidence. Basins were defined according to GLNPO (2003) as listed in Table 1.

Assessment of such a change can be accomplished with a two sample *t*-test. Therefore, determination of the minimum detectable difference currently permitted by the data can be computed using the following formula:

$$\delta = \sqrt{\frac{2s_p^2}{n}}\left(t_{\alpha,\nu} + t_{\beta(1),\nu}\right)$$

where:

s_p^2 = sample estimate of pooled population variance; and

δ = the minimum detectable difference.

It was also necessary to make some assumptions about which variates should be subject to the DQO. In this report, the following major taxonomic groupings were assessed: total cladocerans, total adult cyclopoids, total

Table 2. Sources of variability in zooplankton analysis.

Source of Variability	Current Measure
Within-Basin Spatial Heterogeneity	Replicate Stations Within Basin
Sample Collection, Small-Scale Patchiness	Replicate Field Tows at Master Stations
Sub-Sampling	None
Laboratory Analysis	Duplicate QC Counts on 10% of Samples

cyclopoid copepodites, total adult calanoids, total calanoid copepodites and total crustaceans excluding nauplii. In all cases, density rather than biomass was used. Groups constituting less than 20% of total density for any basin/season combination were excluded from the analysis. In addition, minimum detectable differences were calculated for several of the most common species. These included the cladocerans *Daphnia galeata mendotae* and *Bosmina longirostris*, the cyclopoid copepod *Diacyclops thomasi*, and the calanoid copepods *Leptodiaptomus minutus* and *Leptodiaptomus ashlandi*. Only data generated from the deeper, 153-µm mesh tows were assessed. Estimates of variance were calculated from 1998 data using only regular field samples.

4.3 Sources of variability

There are problems posed in trying to assess the variability of multivariate data. Conventional indices of dispersion, e.g., standard deviation, interquartile range, etc., are strictly speaking not applicable to multivariate data, and therefore if used must be applied either to broad summations of the data (e.g., total numbers of crustaceans, total numbers of cladocerans, etc.), or must be calculated separately for each individual variate (i.e., each taxonomic group). This results in a multitude of estimates of variability for each sample, the exact number of which depends upon the number of species encountered in that sample. The collective interpretation for a given sample of these estimates of variability is problematic.

Alternatively, recourse can be made to multivariate techniques. A number of different numerical techniques have been developed in ecology to quantify degrees of identity between pairs or groups of samples which treat this multivariate data as a whole. Among these techniques, measures of similarity seek to provide objective measures of the degree of identity in the structure of two communities. Typically these indices involve summing up the differences in the abundances or biovolumes of individual species between two samples/sites, which reduces these differences to a single number scaled between 0 and 1. The inverse of these measures, i.e., dissimilarity, can also be computed to quantify the distance of two samples from each other. Where a number of samples are assumed to represent the same 'population' (used here in a statistical sense), then the calculation of a matrix of similarity values between these samples can be used to represent the degree of variability among those replicate samples. While this approach has the dual advantage of treating multivariate data in its entirety, and of reducing comparisons between samples to a single number, the drawbacks are that these techniques, when used as measures of variability, are not strictly comparable with more standard methods, and furthermore, the characteristics (e.g., expected distributions) of the numbers generated by these comparisons are not fully defined, as is the case with, for example, estimates of parametric variance. Also, when more than two samples are compared, the resulting similarity comparisons produce a matrix of values rather than a single value, and thus the necessity of reducing these to a single number remains. Unlike the multiplicity of variances produced by analyzing each variate separately by more conventional means, though, the values of a similarity matrix all estimate the same thing, namely the degree of dispersion amongst a set of replicates. In spite of these drawbacks, the benefits provided by a technique capable of fully comparing sets of multivariate data recommend its use in the present context.

Here, both approaches (i.e., calculation of parametric variance on individual variates and comparison of samples using similarity indices) were used to assess the variability of GLNPO's zooplankton data. While these two methods are complementary, their results are also largely incommensurate, quantitatively, and this should be borne in mind when interpreting the results presented here.

4.3.1 ANOVA analyses

To assess the relative contributions of the various stages of sample collection and analysis outlined in Table 2 to the overall variability of zooplankton data, analyses of variance were conducted. The sample analysis scheme of the zooplankton program can be thought of as being comprised of a number of hierarchical stages. Within each lake, basins have been defined by GLNPO to be statistically homogeneous regions. Within basins, stations serve as replicates. Multiple tows, performed at master stations, in turn serve as subsamples within those stations. Duplicate laboratory analyses, finally, serve as 'subsamples' of sample analysis. The variance associated with each of these hierarchical levels can be estimated using a multi-factor nested analysis of variance (ANOVA). The theoretical factor structure of the GLNPO zooplankton data is illustrated in Fig. 2.

In fact, though, the zooplankton data presents an extremely unbalanced statistical design. Field replicates are only nested within one station per basin (the master station), and duplicate laboratory analyses are conducted, on average, on only one sample per lake, and are rarely nested within field replicates. This both complicates the calculation of the ANOVA, and can also lead to anomalous results. Specifically, an unusually high degree of variability in a single pair of analyses at one level of replication (e.g., laboratory duplicate analysis) can mask the variability in the next higher level of subsampling (e.g., field replication).

ANOVA analyses were carried out on six variates: total adult calanoids, total calanoid copepodites, total adult cyclopoid copepods, total cyclopoid copepodites, total cladocerans, and total crustaceans, exclusive of nauplii. Only data generated from the deeper, 153-µm mesh nets were used. Data were natural log transformed prior to analysis; where zeros occurred in the data, 1 was added to all values prior to transformation. Separate analyses were conducted for the two years examined (1998, 1999) and the two seasons (spring,

<table>
<tr><td colspan="18">BASIN 1</td><td colspan="18">BASIN 2</td></tr>
<tr><td colspan="6">Site 1</td><td colspan="6">Site 2</td><td colspan="6">Site 3</td><td colspan="6">Site 1</td><td colspan="6">Site 2</td><td colspan="6">Site 3</td></tr>
<tr><td colspan="2">FD 1</td><td colspan="2">FD 2</td><td colspan="2">FD 3</td><td colspan="2">FD 1</td><td colspan="2">FD 2</td><td colspan="2">FD 3</td><td colspan="2">FD 1</td><td colspan="2">FD 2</td><td colspan="2">FD 3</td><td colspan="2">FD 1</td><td colspan="2">FD 2</td><td colspan="2">FD 3</td><td colspan="2">FD 1</td><td colspan="2">FD 2</td><td colspan="2">FD 3</td><td colspan="2">FD 1</td><td colspan="2">FD 2</td><td colspan="2">FD 3</td></tr>
<tr><td>X</td><td></td><td></td><td></td><td></td><td></td><td>X</td><td></td><td></td><td></td><td></td><td></td><td>X</td><td></td><td>X</td><td></td><td>X</td><td></td><td>X</td><td>X</td><td></td><td></td><td></td><td></td><td>X</td><td></td><td>X</td><td></td><td>X</td><td></td><td>X</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Fig. 2. Illustration of factor structure for hierarchical analysis of variance of GLNPO zooplankton data for hypothetical two basin lake. FD indicates field replicate; cells for laboratory replicates are

summer). Cladocerans were not analyzed in spring samples due to low numbers. In all a total of 22 analyses were performed.

Sources of variation included between basin variance, between station within basin), between field replicate within station) and between laboratory duplicate (within field replicate) variance. The structure of the analysis assumed that the amount of variance contributed by each factor was similar for all levels of that factor, so that, for instance, between station variability was similar within all basins. However, it was noted that the variability between stations in the western and central basins of Lake Erie was extremely high. In order that this not exert an undue influence, these two basins were removed from the analysis. The magnitude of the different variance components was computed as a percentage of the total variance minus between basin variance, i. e., variance components were calculated as a percentage of *within* basin variance.

This approach can provide information about the amount of variability involved in estimating densities of major taxonomic groups. However, it cannot address variability in estimates of species composition. This distinction should be borne in mind when interpreting the results. If the species composition of the zooplankton community within a basin is consistent from site to site, but the total numbers of organisms vary widely, an ANOVA will indicate high levels of variability. On the other hand, if the species composition of the community is vastly different from site to site, but densities of individuals are similar within each broad taxonomic category, then an ANOVA will indicate low variability.

4.3.2 Similarity analyses

As indicated earlier, special problems are posed in trying to quantify the variability of multivariate data. While the data can be summarized by broad taxonomic category into a smaller number of individual variates, and variance calculated using univariate methods as outlined above, this approach will not be able to detect compositional shifts at lower taxonomic levels, and thus cannot give a true picture of variability at the community level. It is desirable, instead, to use a measure of variability that can simultaneously compare all the variates within samples, and which can produce a single number to quantify the degree to which the samples diverge.

The approach adopted here involves measures of similarity/dissimilarity. These measures compare two multivariate samples and produce a single number indicating to what extent the two samples share the same species, and optionally to what extent those species are present in similar densities in the two samples. It is important to bear in mind that a similarity value is the result of a comparison between *two samples*. To compare a set of replicates, then, each replicate must be compared with each other replicate, and a matrix of similarity values obtained, from which some measure of central tendency (e.g., median, mean) can be computed. Thus for N samples, [N(N-1)]/2 comparisons would be performed.

The primary differences between most similarity indices have to do with whether each species will be compared on the basis of presence/absence, relative abundance, or absolute abundance. Where relative abundances are compared, the similarity measure will be sensitive to differences in species composition, but not to variability associated with estimating overall densities. Where absolute abundances are used, variability in both species composition and densities will be quantified with the similarity measure. Using both types of similarity measures in tandem, therefore, provides a means of assessing whether the variability between two samples is due primarily to differences in species composition, or differences in densities.

Of the similarity measures based on comparisons of relative abundances, one of the most intuitive and most commonly used is the Percentage Similarity of Community (PSc) index of Whittaker (1952; Whittaker and Fairbanks, 1958). As suggested by its name, this index compares percent abundances of species in two samples. Therefore, if two samples have vastly differing total numbers of individuals, but the species within each sample contribute exactly the same proportion of individuals, then the PSc index will indicate that the two samples are identical. The index is calculated as:

where *a* and *b* are, for a given species, the relative proportions of the total samples *A* and *B*,

$$PS_c = 1 - 0.5\sum_{i=1}^{K}\left|a-b\right|$$

respectively, which that species represents. The absolute value of their difference is then summed over all *K* species. Two samples in which all species are present in identical proportions will result in a score of 1 (or 100%), while two samples sharing no species in common will produce a score of 0.

Another widely used index, but one which compares absolute abundances of species in two samples, is the so-called Bray-Curtis index. Originally developed by Kulczyński (1927), and subsequently modified by Motyka *et al.* (1950), this index provides a number from 0 (no species in common) to 1.0 (identical samples) similar to that of Whittaker's PSc index. The index is calculated as:

where a = the sum of all species abundances

$$C = 2\frac{W}{a+b}$$

in sample in sample, b = the sum of all species abundances in the other sample, W = the smaller of the two abundances for each species, summed over all species. In this report, this index will be referred to as C, in accordance with its presentation in Motyka *et al.* (1950). When these two indices are used together, they can provide both qualitative (i.e., relative) and quantitative information about the similarity of two samples. Specifically, when C values are substantially lower than PSc values, this indicates that differences between the two samples derive at least in part from differences in absolute numbers of individuals in the two samples. Where the two values are substantially the same, then differences between the two samples are due primarily to differences in species composition.

To quantify levels of variability associated with natural variation and different sample collection/analysis activities, similarity matrices were computed between samples taken within each basin (separated by season and mesh size), between sets of field replicates, and between duplicate laboratory analyses. Separate matrices were generated for spring and summer, and 63- and 153-µm mesh tows.

Differences in similarity values generated from the two different measures, as well as differences in values from each measure due to season and mesh size, were assessed using a Mann Whitney rank sum test. While it would have been preferable to use a multifactor ANOVA to assess all factors simultaneously, no transformation was found that could stabilize variance and ameliorate the non-normality of the data, and formulations for a non-parametric, multifactor ANOVA type test could not be found.

To estimate the relative contributions of within basin spatial heterogeneity, sample collection, and laboratory analysis to the variability of the data, similarity values were converted to dissimilarity values by subtracting them from 1. To determine the relative magnitudes of each source of uncertainty, the mean dissimilarity associated with each stage was subtracted from that of the previous stage. For example, to determine the amount of dissimilarity contributed by sample collection, the dissimilarity estimates generated from QC analyses were subtracted from those generated from field replicates. Likewise, an estimate of the amount of dissimilarity contributed by site to site variability was obtained by subtracting the dissimilarity of field replicates from within-basin dissimilarity values.

5 Results

5.1 Minimum detectable differences

The percent minimum detectable differences for total crustaceans ranged between 31% (southern basin of Lake Michigan, spring) and 176% (western basin of Lake Erie, spring), with a median of 63% (Fig. 3). For this response variable, no basin/season met the DQO. The highest values were seen in Lake Erie, although all lakes had at least one value approaching or exceeding 100%. For these basin/seasons, therefore, the current sampling

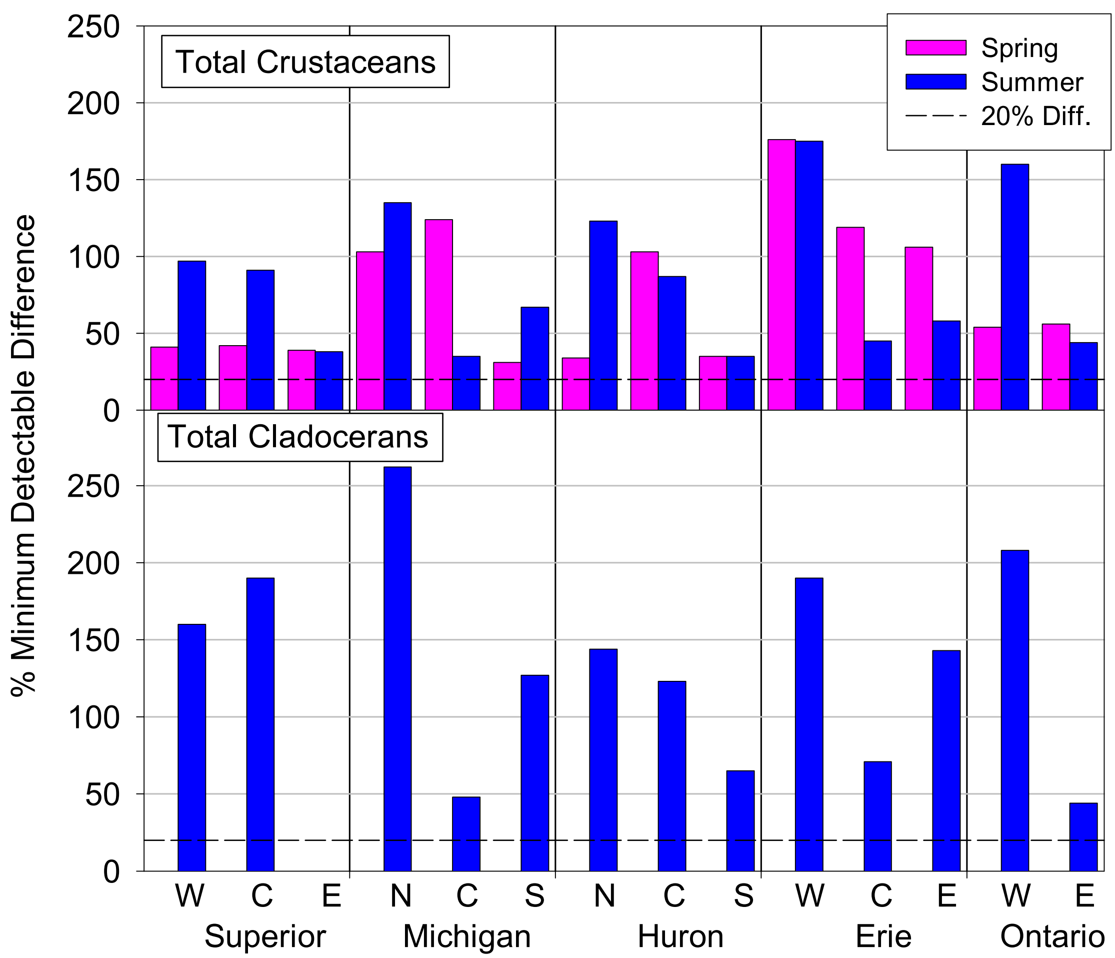

Fig. 3. Percent minimum detectable differences for total crustaceans and total cladocerans. Variances calculated from 1998 data.

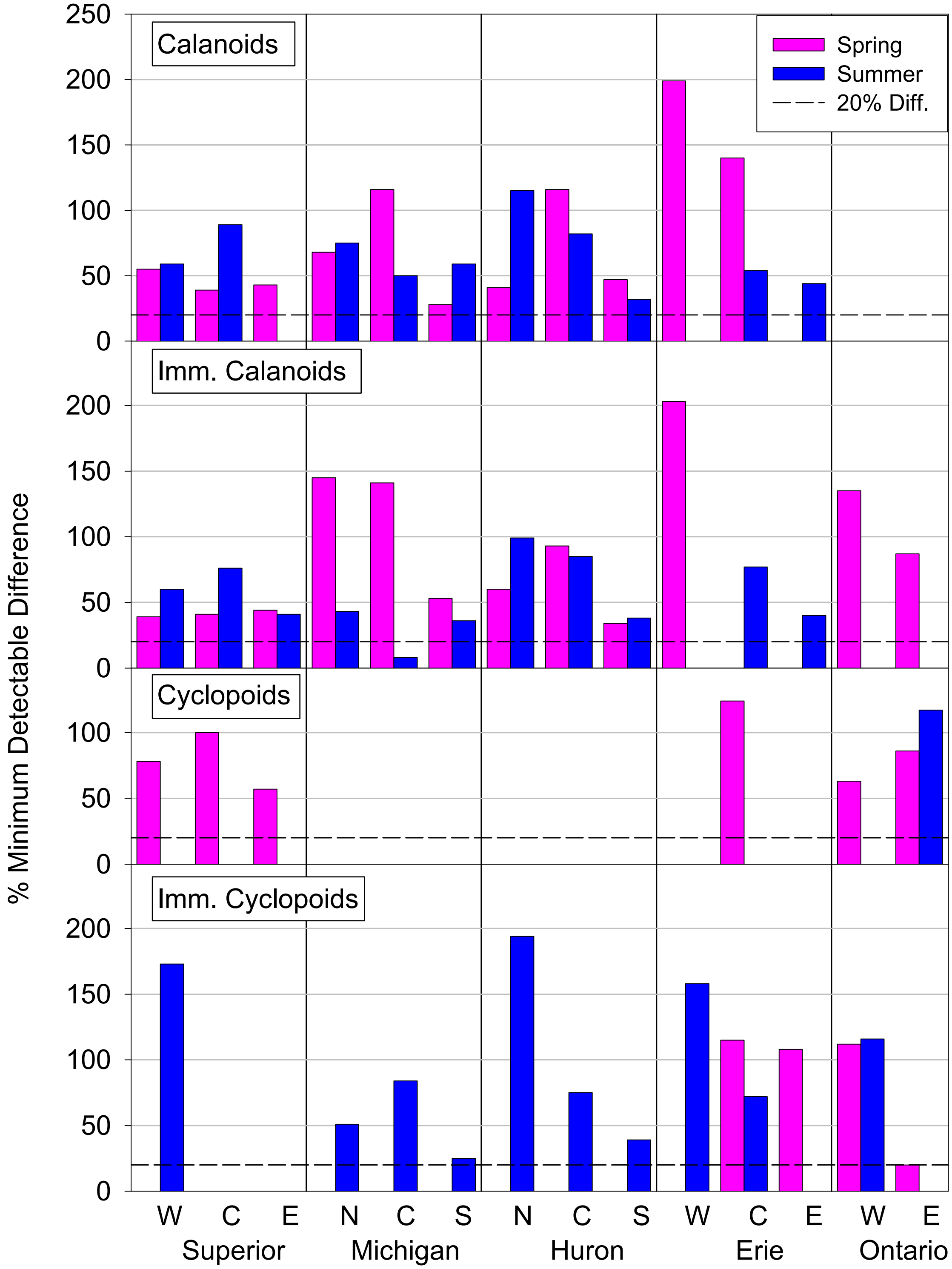

Fig. 4. Percent minimum detectable differences for adult calanoid copepods, immature (copepodite) calanoid copepods, cyclopoid copepods and immature (copepodite) cyclopoid copepods.

regime would have an 80% chance of detecting a change in total crustacean density with 90% confidence only if that change constituted at least a doubling in density. Percent minimum detectable differences for total cladocerans could only be assessed for summer samples, due to low numbers in spring samples. These were substantially higher than for total crustaceans, with basin-wide values ranging from 44% (eastern basin of Lake Ontario) to 262% (northern basin of Lake Michigan), and an overall median value of 143% (Fig. 3). Again, no basin met the DQO requirements.

Minimum detectable differences were lower for both adult and immature calanoid copepods (Fig. 4), and this was probably due at least in part to the great numbers of these individuals found at most sites. Median percent minimum detectable differences were 59% and 60%, respectively, for these groups. Percent minimum detectable differences for cyclopoids were intermediate between cladocerans and calanoids, again probably due in part to their relative abundances (Fig. 4). Median percent minimum detectable differences for adult and immature cyclopoids were 86% and 96%, respectively. Among the copepod groups, the DQO was met in only two cases: calanoid immatures in the central basin of Lake Michigan in the summer and cyclopoid immatures in the eastern basin of Lake Ontario in the spring. Overall, percent minimum

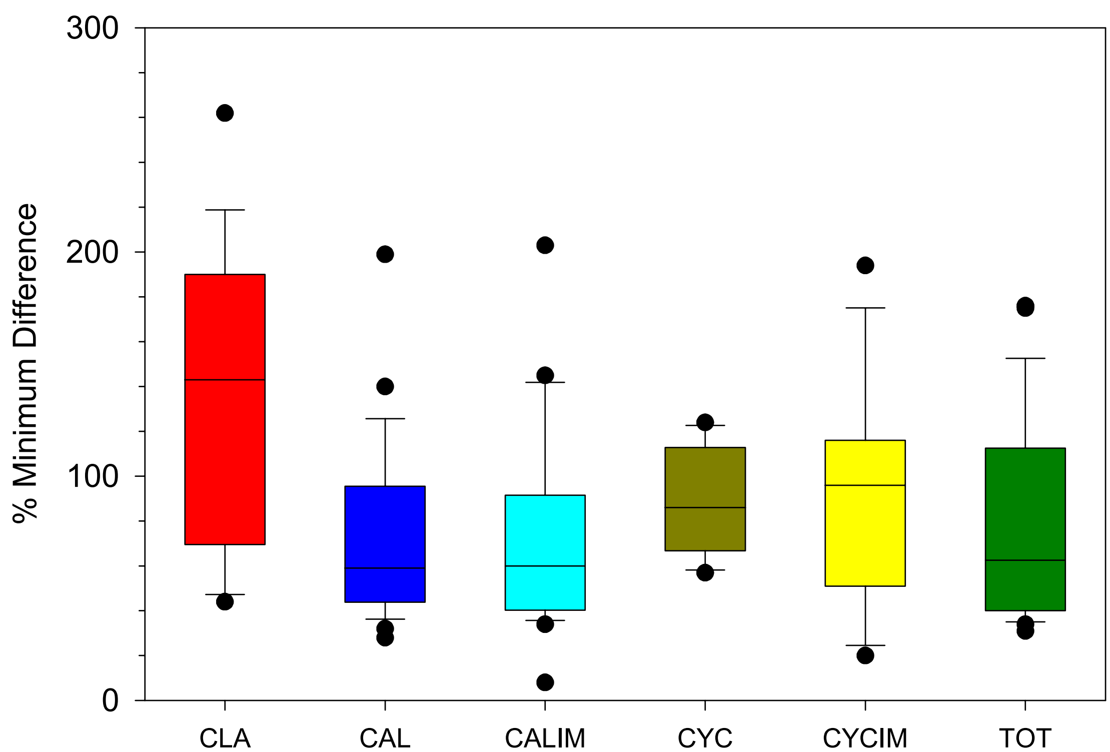

Fig. 5. Percent minimum detectable differences for major taxonomic groups CLA = total cladocerans; CAL= total adult calanoids; CALIM = total calanoid copepodites; CYC = total adult cyclopoids; CYCIM= total cyclopoid copepodites; TOT = total crustaceans, exclusive of nauplii. Boxes indicate 25th and 75th percentiles; whiskers denote 10th and 90th percentiles; lines denote median; symbols denote outliers.

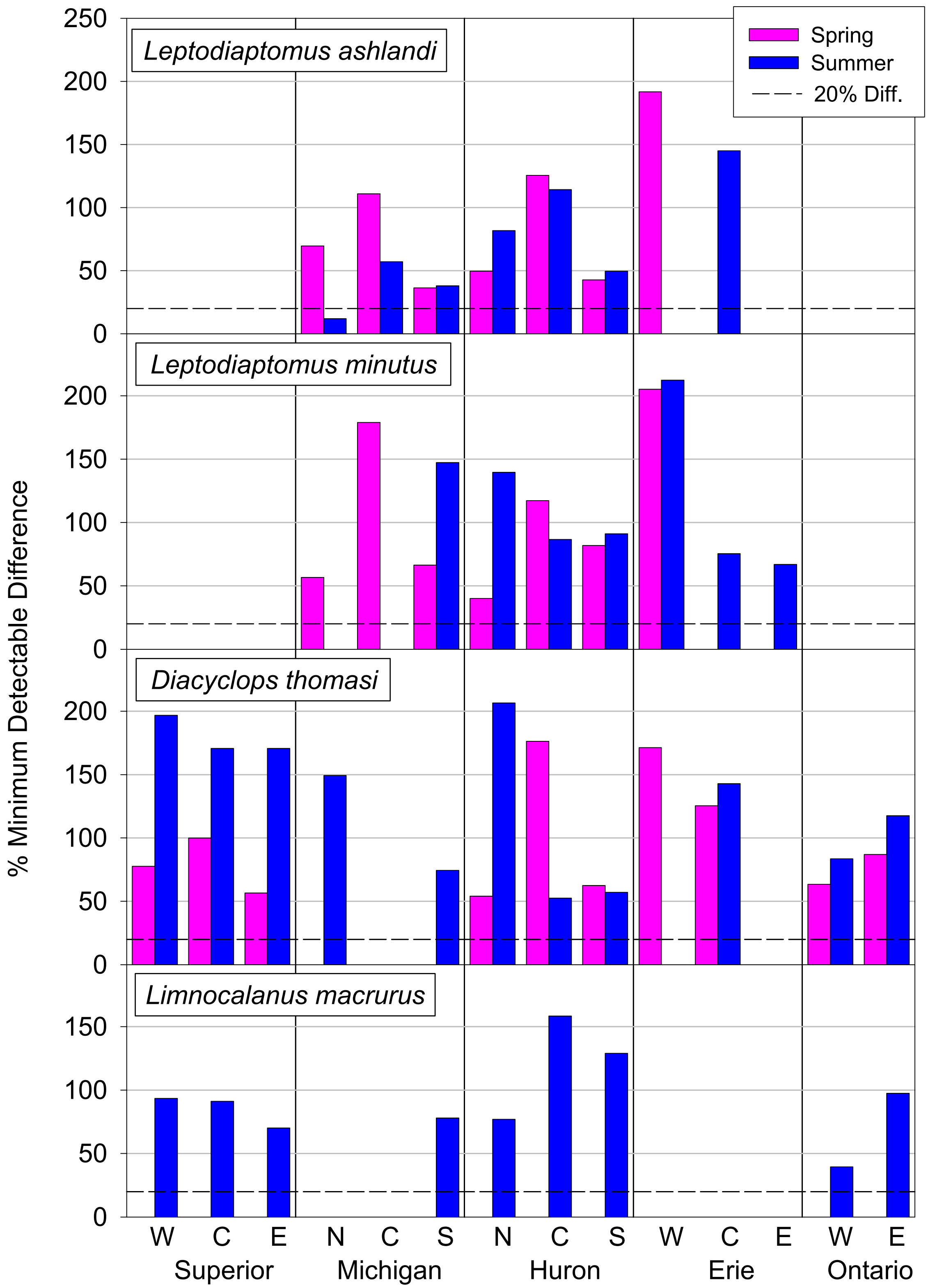

Fig. 6. Percent minimum detectable differences for the most common adult calanoid copepod species.

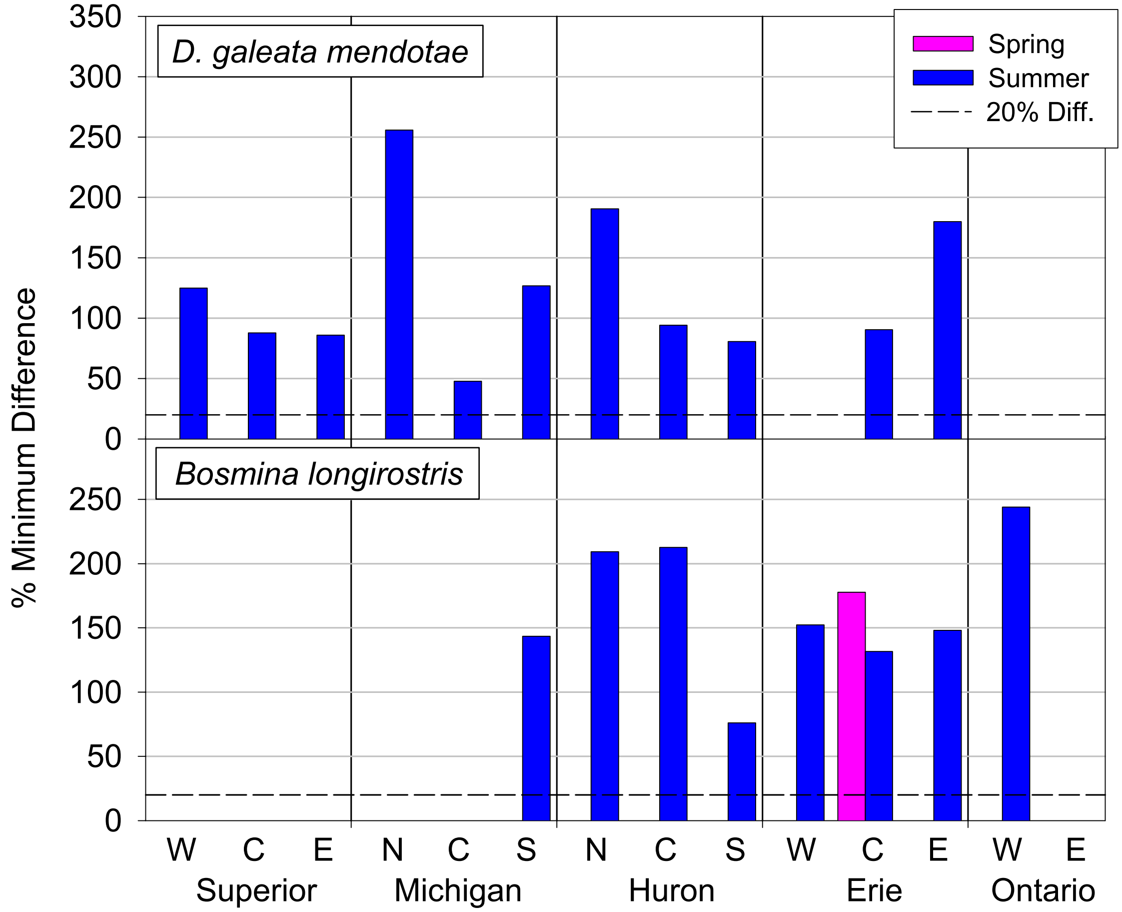

Fig. 7. Percent minimum detectable differences for the most common cladoceran species.

detectable differences were highest for cladocerans, lowest for calanoids, and intermediate for cyclopoids (Fig. 5).

Of the six individual species examined, in only one case was the DQO requirement met (*L. ashlandi*, Lake Michigan, northern basin, summer). Percent minimum detectable differences ranged from 12% to 256%, with an overall median of 94% (Figs 6 and 7). This suggests that, on average, the density of a species would have to double from one year to another in order for the current sampling regime to be able to detect the change as statistically significant. Overall, the two cladocerans (*D. galeata mendotae* and *B. longirostris*) had higher percent minimum detectable differences than the copepods examined. As with the larger taxonomic groupings, there were no clear lake to lake differences in percent minimum detectable differences for the individual species.

When considered in aggregate on the basis of lake basin, minimum detectable differences were consistently higher in the western basin of Lake Erie than in the other basins (Fig. 8). The eastern basin of Lake Superior and the southern basin of Lake Huron exhibited consistently low minimum detectable differences.

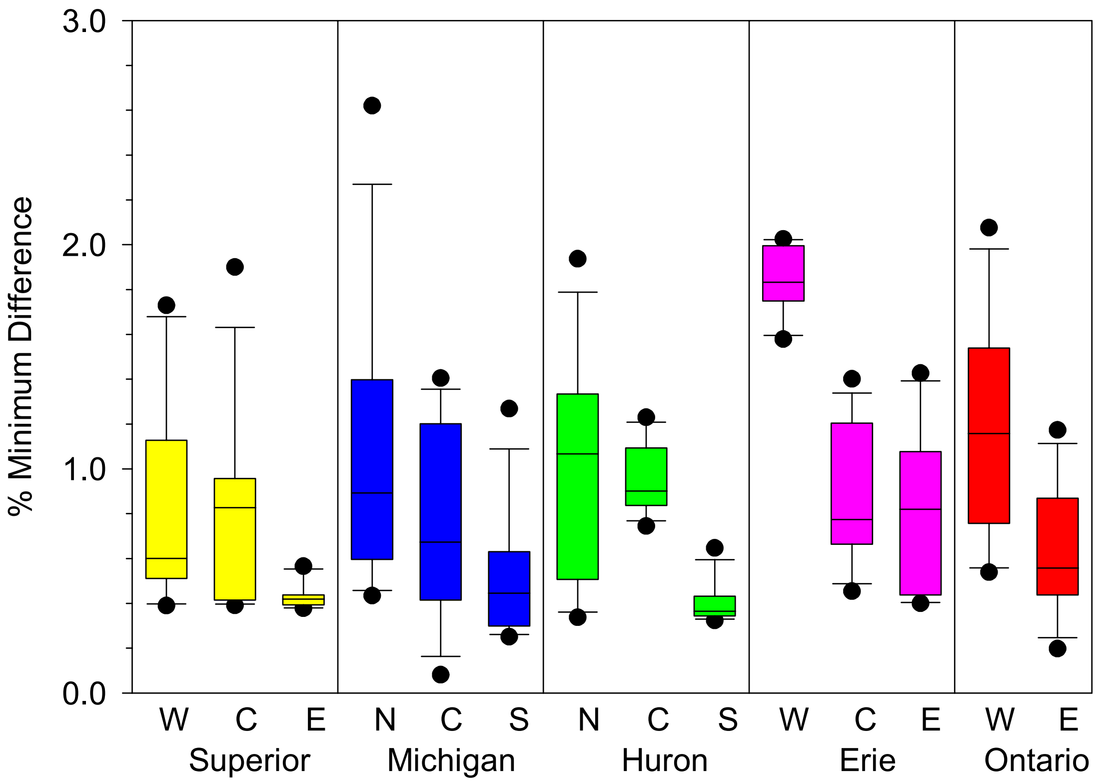

Fig. 8. Percent minimum detectable differences basins. Box plots as in Fig. 5.

Aside from these instances, though, percent minimum detectable differences were highly variable, and clear basin-to-basin differences were not seen.

5.2 Sources of variability of zooplankton data - ANOVA analyses

In almost all cases, the largest source of variance in the estimation of within-basin abundances of major taxonomic groups was associated with between-station variability (Table 3). This contributed from 23% (summer, 1999, total crustaceans) to nearly 95% (summer, 1998, adult cyclopoids) of the within-basin variance. On average, between-station variance made up about 70% of the total within basin variance. This suggests that large scale spatial heterogeneity in abundances is the main source of uncertainty in developing basin-wide estimates of crustacean abundances.

Variances associated with field replicates contributed on average 26% to total within-basin variance, and ranged from 2.4% (summer 1998, cyclopoids) to 76.7% (summer 1999, total crustaceans). It should be borne in mind that since replicates are not taken at the point of subsampling in the laboratory, variances calculated from field replicates would also incorporate that variance component. The least amount of variance was contributed by duplicate laboratory (QC) analyses, which on average contributed less than 4% of total within-basin variance. Relatively few duplicate

Table 3. Relative contributions of different sources of variance to the estimation of within-basin abundances of major taxonomic grouping, as determined by multi-stage hierarchical ANOVA.

Variance Comp	Cal	Cal Imm	Cla	Cyc	Cyc Imm	Total
Spring 1998						
Between Station	42.8%	80.7%		81.5%	76.2%	83.6%
Field Reps	47.9%	17.8%		17.9%	15.6%	16.1%
Lab Dups	9.3%	1.5%		0.6%	8.2%	0.4%
Summer 1998						
Between Station	48.7%	51.8%	80.7%	94.9%	87.1%	70.3%
Field Reps	50.0%	19.8%	19.2%	2.4%	0.0%	25.3%
Lab Dups	1.3%	28.4%	0.1%	2.7%	12.9%	4.3%
Spring 1999						
Between Station	69.3%	78.6%		79.6%	79.2%	78.5%
Field Reps	28.0%	20.1%		19.7%	17.0%	21.3%
Lab Dups	2.7%	1.3%		0.7%	3.9%	0.1%
Summer 1999						
Between Station	64.8%	41.6%	82.5%	72.8%	79.0%	23.2%
Field Reps	34.9%	58.2%	17.5%	26.3%	20.1%	76.7%
Lab Dups	0.4%	0.2%	0.0%	0.8%	0.9%	0.1%

Cal = total adult calanoids; Cal Imm = total calanoid copepodites; Cla = total cladocerans; Cyc = total adult cyclopoids; Cyc Imm = total cyclopoid copepodites; Total = total crustaceans, exclusive of nauplii.

laboratory analyses are carried out, so a single aberrant counts can have a large impact on this analysis. This was the case in Summer, 1998, when one set of duplicate laboratory analyses from Lake Ontario yielded highly divergent estimates of immature copepod densities. This resulted in both unusually inflated variance estimates for laboratory duplicates for immature calanoids and immature cyclopoids, and anomalously low error estimates of field replicate variance for those two variates.

In summary, then, it appears that the majority of uncertainty involved in the estimation of crustacean abundances, at least viewed at the level of order and suborder, results from large scale (i.e., station to station) spatial heterogeneity, while a relatively minor amount is due to inaccuracies in counting on the part of laboratory analysts. Somewhat less than one third comes from errors associated with sample collection and/or subsampling in the laboratory. Given the broad taxonomic groupings used in this analysis, error due to taxonomic inaccuracies would not be included in these estimates of variance.

5.3 Sources of variability of zooplankton data - similarity analyses

5.3.1 Duplicate laboratory (QC) analyses

A total of 74 sets of duplicate laboratory (QC) analyses from both 63-µm and 153-µm mesh nets, and both spring and summer cruises, were assessed, using both PSc and C similarity indices. Data were from 1998 and 1999, the only two years for which full datasets of 153-µm mesh net tows are currently available. It will be recalled that these values quantify the similarity between tabulated species composition estimates generated by two different analysts counting the same sample, subsequent to sample splitting. Similarity values using both measures were uniformly high (Table 4); 95% of PSc values were above 0.91, while 95% of C values were above 0.88. Median values for both measures were 0.97. Statistically significant differences ($\alpha = 0.05$) between lakes, mesh size or season were not found, which suggests that taxonomic difficulties are not more marked in any given lake or season, or for shallow or deeper tows. Differences between similarity values calculated using PSc and C also were not apparent. Such differences would arise from discrepancies in absolute counts of organisms, and the absence of differences between the two measures indicates that analysts have little trouble consistently counting all of the organisms in the counting chamber, a conclusion also supported by the ANOVA results. Subsamples are chosen specifically to ensure a relatively narrow range of individual organisms in the counting chambers - generally between 200 and 400 - so large discrepancies in counts of individuals would not be expected.

QC limits have as yet not been agreed upon for zooplankton analyses. Based on the present analysis, if duplicate QC counts are compared using the PSc index, a value of 0.92 should be expected in 90% of cases. It is therefore suggested that this value be adopted as a QC limit. This limit should be applicable to both 63-µm and 153-µm mesh tows taken during both spring and summer. QC criteria based on PSc values would guard against taxonomic errors, but not enumeration errors.

When all QC analyses from 1998 and 1999 are examined, the majority of discrepancies between total counts of organisms resulting from duplicate laboratory analyses are less than 2% of the average of the two counts

Table 4. Percentiles of Whittaker PSc and C similarity values for comparisons between pairs of duplicate laboratory (QC) analyses. Data were from 1998 and 1999, and include data from both spring and summer cruises and both 63- and 153-µm mesh net tows.

PSc		C	
Percentile	Similarity	Percentile	Similarity
95th	0.99	95th	0.99
75th	0.98	75th	0.98
50th	0.97	50th	0.97
25th	0.96	25th	0.95
5th	0.91	5th	0.88

Table 5. Percentiles of relative discrepancies in counts of total organisms between duplicate laboratory analyses. Relative discrepancies (Δ Count %) are calculated as [absolute(count#1-count#2)/ average(count#1, count#2)]*100. Data is from 1998 and 1999, and includes both spring and summer samples, as well as 63-µm and 153-µm mesh tows.

Percentile	Δ Count (%)
95th	5.80%
75th	2.69%
50th	1.53%
25th	0.57%
5th	0.10%

(Table 5). In 90% of cases, differences between duplicate counts amounted to just over 4% of the average of the two counts. It is therefore recommended that a relative percent difference of 4% be adopted as a criterion for total organism counts of duplicate QC analyses, with those analyses exceeding this limit subject to recounts by both analysts.

5.3.2 Field replicates

PSc similarity values between field replicates were on average quite high, with 90% of all values ranging between 0.84 and 0.97, and an overall median of 0.93 (Table 6, Fig. 9). This range is not dramatically lower than similarity values of QC samples, and indicates that relatively little variability is introduced during the sampling process as far as relative proportions of taxa are concerned. PSc similarity between field replicates taken during the summer cruises was slightly lower than similarity of spring field replicates, and this difference, though slight, was statistically significant (Table 7). No systematic differences were found between tows using different mesh sizes (i.e., deep and shallow tows) (Table 8).

Table 6. Percentiles of Whittaker PSc and C similarity values for comparisons between field replicate analyses. Data were from 1998 and 1999, and include both spring and summer cruises and both 63- and 153-µm mesh net tows.

PSc		C	
Percentile	Similarity	Percentile	Similarity
95th	0.97	95th	0.95
75th	0.95	75th	0.91
50th	0.93	50th	0.86
25th	0.90	25th	0.78
5th	0.84	5th	0.63

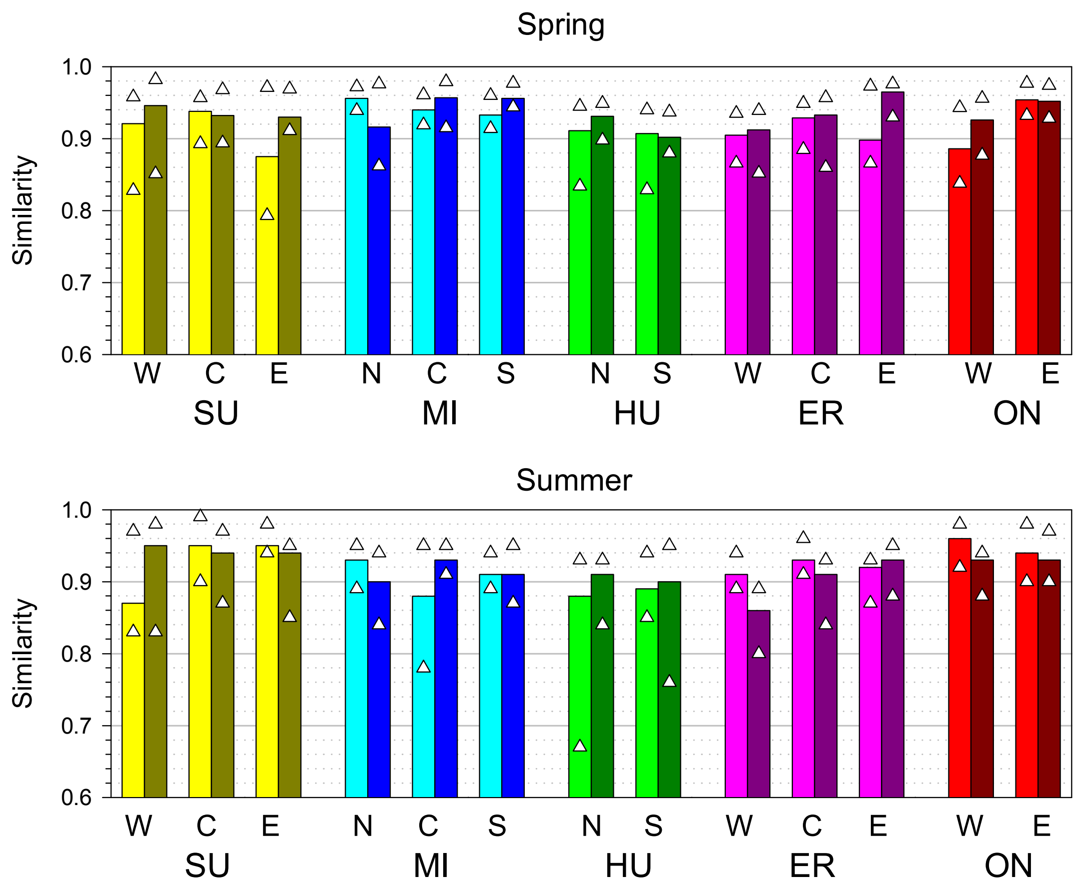

Fig. 9. PSc similarity values between field replicates collected in 1998 and 1999. Bars indicate means; triangles indicate minimum and maximum values for each set of comparisons. Comparisons between 63-µm mesh tows are left (lighter) bars, comparisons between 153-µm tows are right (darker) bars).

Table 7. Results of Mann Whitney rank sum test comparing effects of season on values of PSc similarity comparisons between field replicates.

Group	Median	25%	75%
Spring	0.935	0.900	0.950
Summer	0.920	0.890	0.940

T = 26492.0 P = 0.009

Table 8. Results of Mann Whitney rank sum test comparing effects of mesh size on values of PSc similarity comparisons between field replicates

Group	Median	25%	75%
153 µm	0.928	0.899	0.949
63 µm	0.928	0.894	0.943

T = 25041.0 P = 0.432

Similarity between field replicates calculated using the C index were substantially lower than PSc values (Fig. 10, Table 6); this difference was highly statistically significant (Table 9). C values also exhibited a broader range than PSc values, and in particular contained more extremely low values. The difference in similarity values calculated by the two indices indicates that field replicates are more variable in their estimates of zooplankton densities, while being relatively consistent in their estimates of percent contributions of individual species.

Table 9. Results of Mann Whitney rank sum test between PSc and C similarity values.

Group	Median	25%	75%
C	0860	0.780	0.910
PSc	0.930	0.900	0.950

T = 69913.0 P = <0.001

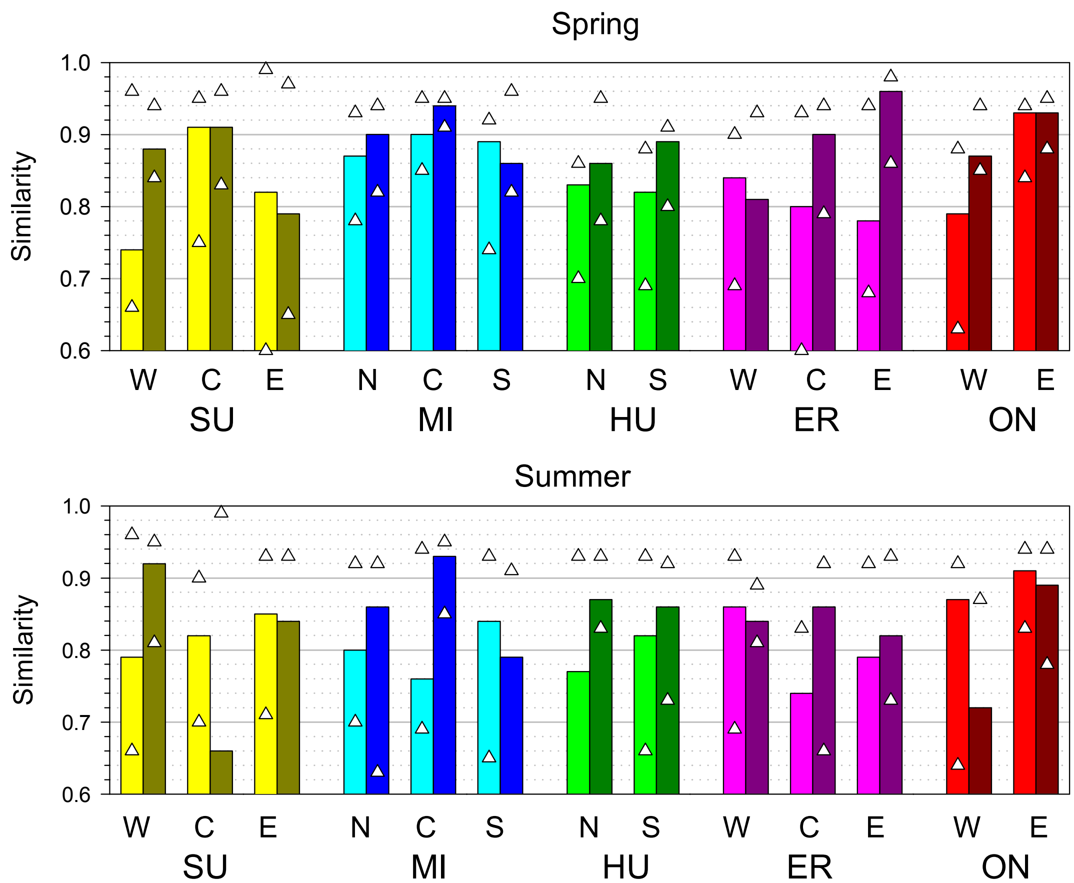

Fig. 10. C similarity values between field replicates collected in 1998 and 1999. Bars indicate means; triangles indicate minimum and maximum values for each set of comparisons. Comparisons between 63-μm mesh tows are left (lighter) bars, comparisons between 153-μm tows are right (darker) bars.

Variability in the estimation of densities between field replicates can result from a number of possible factors. Zooplankton patchiness on the spatial scale of the replicate tows - a scale dependent upon how much the vessel drifts between replicate tows - would introduce variability into density estimates. Variability could also result from inaccuracies in flowmeter readings, due either to malfunction or to misreading on the part of the technician, or it can be due to differences between replicates in the angles at which the net is towed. To test these last two possibilities, regressions were run between the minimum C index values within each set of field replicate comparisons and the maximum angle of the net for those field replicates, the maximum difference in net angle among the field replicates, the maximum relative difference in flowmeter readings amongst the three field replicates, and the depth specific maximum relative difference in flowmeter readings amongst the three field replicates. These latter two independent variables were calculated as follows:

$$\frac{\max \mathit{flowmeter} - \min \mathit{flowmeter}}{\max \mathit{flowmeter} + \min \mathit{flowmeter}}$$

$$\frac{\max \mathit{flowmeter}}{\mathit{depth}} - \frac{\min \mathit{flowmeter}}{\mathit{depth}}$$

Prior to analysis, C values were transformed using an arcsin square root transformation to normalize the data. After transformation, data met assumptions of normality and homoscedasticity.

No relationship was found between C values and net angle. However, a highly significant relationship was found between C values and differences in flowmeter readings between field replicates (Table 10). This relationship explained slightly less than a third of the variance in C values. A similar relationship was found when depth specific flowmeter values were examined. Therefore, it appears that a portion of the variability involved in sample collection is due to inconsistencies in flowmeter readings amongst the field replicates. As noted, this could result from variability in the meter itself, or from inconsistencies in reading the meter.

The majority of variance in C values, however, was not accounted for by flowmeter readings. This points to patchiness of zooplankton populations, other aspects of sample handling, such as washing the net, decanting into bottles, etc., or variance associated with

Table 10. Regression results of C values and maximum relative difference in flow meter readings between field replicates.

Arcsin sqrt(B-C) = 1.166 - (0.380 * Relative diff. in flowmeter readings)

	Coefficient	Std. Error	t	P
Constant	1.166	0.0186	62.7	<0.001
Rel diff flow	-0.380	0.0555	-6.8	<0.001

Analysis of Variance:

	DF	SS	MS	F	P
Regression	1	0.837	0.837	46.8	<0.001
Residual	102	1.823	0.0179		
Total	103	2.661	0.0258		

N = 104 Adj R^2 = 0.308

Table 11. Results of Mann Whitney rank sum test comparing effects of season on values of C similarity comparisons between field replicates.

Group	Median	25%	75%
Spring	0.870	0.810	0.925
Summer	0.840	0.760	0.900

T = 27036.0   P = 0.001

Table 12. Results of Mann Whitney rank sum test comparing effects of mesh size on values of C similarity comparisons between field replicates.

Group	Median	25%	75%
63 µm	0.830	0.755	0.895
153 µm	0.880	0.820	0.920

T = 21390.0   P = <0.001

subsampling in the laboratory as potential major sources of variance for this stage of the analysis.

As with PSc values, there was a significant, though somewhat slight, difference between C similarity values generated from spring and summer cruises, with the latter slightly lower on average than the former (Table 12). This was probably due at least in part to the greater species diversity seen in the summer. A significant difference was also found between mesh sizes, with the smaller mesh size (i.e., shallower tows) showing somewhat greater variability between field replicates, as measured by the C index (Table 12). This difference, though, was not entirely consistent across all basins.

5.3.3 Between station

Within-basin similarity values were only calculated from samples collected with the deeper, 153-µm mesh tows. These similarity values should theoretically provide an estimate of the

Table 13. Percentiles of similarity values for within-basin samples

	PSc		C	
	Percentile	Similarity	Percentile	Similarity
Total	95th	0.93	95th	0.90
	75th	0.87	75th	0.80
	50th	0.79	50th	0.69
	25th	0.69	25th	0.54
	5th	0.47	5th	0.25
Spring	95th	0.94	95th	0.92
	75th	0.90	75th	0.83
	50th	0.85	50th	0.74
	25th	0.76	25th	0.60
	5th	0.58	5th	0.22
Summer	95th	0.89	95th	0.84
	75th	0.83	75th	0.75
	50th	0.74	50th	0.64
	25th	0.61	25th	0.51
	5th	0.36	5th	0.27

Table 14. Results of Mann Whitney rank sum test comparing PSc and C similarity values.

Group	Median	25%	75%
BC	0.690	0.540	0.800
PSc	0.790	0.690	0.872

T = 416276.0 P = <0.001

error contributed to the data from within-basin spatial heterogeneity (in addition to sample collection and analysis). However, the shallower 63-µm mesh tows would also include variation due to vertical migration, since it is likely that some stations within a basin would be visited at different times during the diurnal cycle of at least some species. In order not to confound these two potential sources of variation, therefore, only deeper tows were considered.

As with the field replicate samples, within basin PSc similarity values were higher than C values (Table 13); this difference was statistically significant (Table 14). The differences between these two measures were more pronounced for within basin comparisons than for the field replicates, suggesting that, as might be expected, differences in crustacean densities varied more from site to site than within a site. Half of PSc values were between 0.69 and 0.87, while half of Bray Curtis values ranged between 0.54 and 0.80.

Table 15. Results of Mann Whitney rank sum test comparing effects of season on values of PSc similarity comparisons between field replicates

Group	Median	25%	75%
Spring	0.850	0.761	0.900
Summer	0.740	0.610	0.828

T = 157938.0 P = <0.001

Table 16. Results of Mann Whitney rank sum test comparing effects of season on values of Bray-Curtis similarity comparisons between field replicates.

Group	Median	25%	75%
Spring	0.740	0.600	0.835
Summer	0.645	0.510	0.750

T = 144516.0 P = <0.001

For both measures, similarity values of comparisons made during the spring were statistically significantly higher than those of summer comparisons (Tables 15 and 16). During the spring, over 75% of spring PSc values were above 0.75, a value often taken to indicate samples taken from the same community. Fully half were above 0.85. In contrast, less than half of summer PSc values met the 0.75 criterion. The high values in the spring are most likely reflective of the extremely limited species composition of spring samples. For example, average numbers of crustacean taxa per site ranged between 5 and 8 for the five lakes in spring, 1999. C values were lower than PSc values for both seasons (Table 13). Somewhat less than half of spring values were above 0.75, while only a quarter of summer values met or exceeded that value. The differences between the two indices were more pronounced in spring than in summer, which again indicates that within-basin species composition was more variable in summer.

Values in the central and western basin of Lake Erie were notably lower than those for other basins, and this was apparent for both PSc and Bray Curtis values, indicating that both species composition and densities varied greatly within these two basins (Figs 11 and 12). Consistent differences were not apparent in other basins.

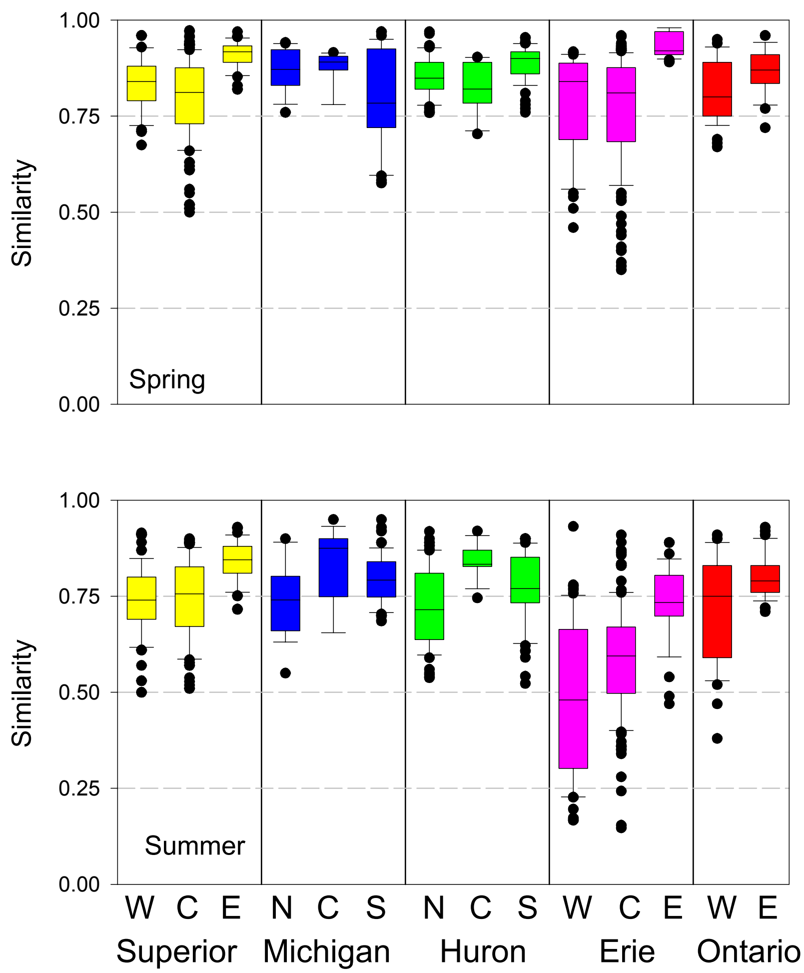

Fig. 11. PSc similarity values for within-basin comparisons. Data from 1998 and 1999; 153-µm mesh tows. Boxes as in Fig. 6.

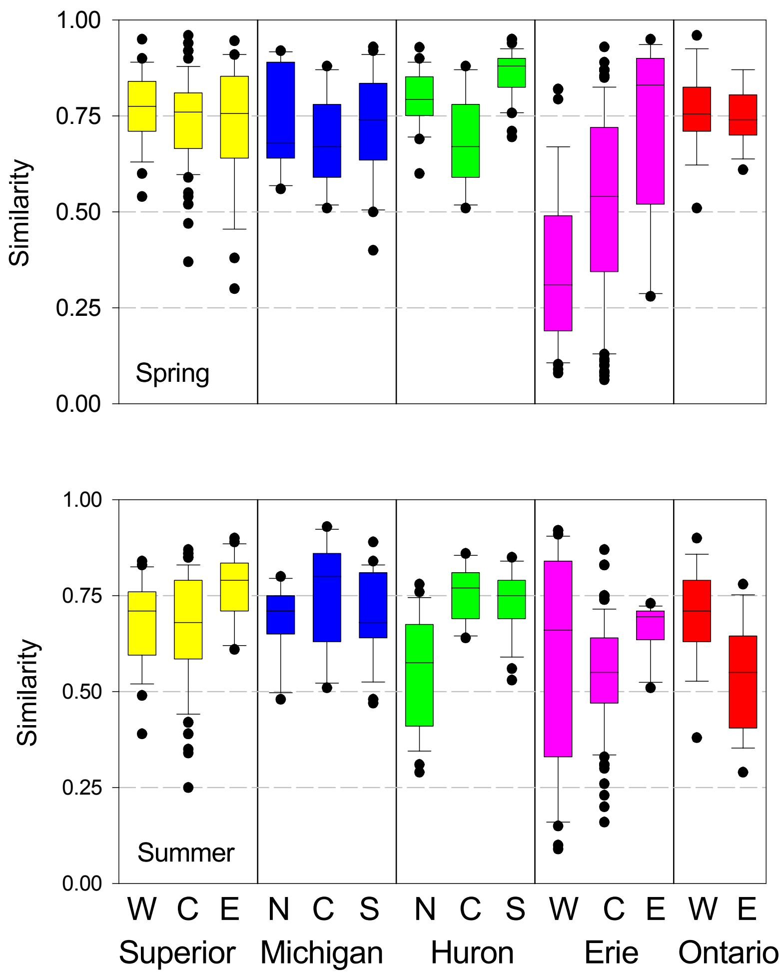

Fig. 12. C similarity values for within-basin comparisons. Data from 1998 and 1999; 153-µm mesh tows. Boxes as in Fig. 6.

5.3.4 Relative contribution of different sources of error

An idea of the relative contribution of the various sources of error can be gained by comparing PSc and C values from the various stages of the analysis. Since what is of interest here is variability, it is more convenient to express these values as dissimilarity values, rather than similarity values. This is accomplished by simply taking their inverse (i.e., 1-PSc; 1-C).

As noted, the amount of uncertainty resulting from laboratory analyses is relatively slight. As measured by dissimilarity this averaged 0.03 (i.e., 1-0.97) for comparisons of spring samples made by both indices, and 0.04 for summer samples (Figs 13 and 14). Values were essentially the same whether relative species composition or actual densities are considered (i.e., when examining PSc or C values). The amount of dissimilarity resulting from sample collection is only slightly higher than that from sample analysis when relative species composition is considered. In other words, estimates of relative species composition appear to be fairly robust for each particular site. Again, it is important to remember that variability due to sub-sampling is not captured by replicate QC analyses, and would therefore be incorporated into dissimilarity values from field replicates. When considered in terms of absolute abundances, however, the contribution of sampling variability increases notably. During spring, on average, it is over three times higher than that of sample analysis, while in summer it is three and a half times greater than that of sample analysis (Fig. 14). This indicates that the greatest introduction of variability during sample collection is in estimation of absolute densities of organisms, while estimates of the relative proportions of constituent species are relatively robust.

In all cases the greatest amount of dissimilarity was a result of site to site variation (Table 17). Even when just relative abundances are compared, site to site variation contributes more dissimilarity than both laboratory analysis and sample collection combined in spring, while this contribution is close to double that of laboratory analysis plus sample collection in summer. When absolute densities (i.e., C values) are considered, the contribution of site to site variation to dissimilarity doubles in spring, but during summer is essentially the same as that of relative proportions of species, indicating that there are substantial differences in species composition from site to site within a basin during the summer, while during spring the majority of dissimilarity results from site to site differences in densities. In all cases, though, dissimilarity values were lower when measuring using the PSc index. The relatively low site to site variability in species composition in spring is consistent with the highly restricted species richness of most spring communities. As was pointed out previously, site to site variability was particularly high in the western and central basins of Lake

Table 17. Relative contributions of different sources of variability to overall within-basin dissimilarity, as measured by both PSc and C indices.

	PSc		C	
Variance Component	Spring	Summer	Spring	Summer
Between Station	51%	64%	54%	45%
Field Reps	27%	19%	35%	42%
Lab Dups	23%	17%	10%	14%

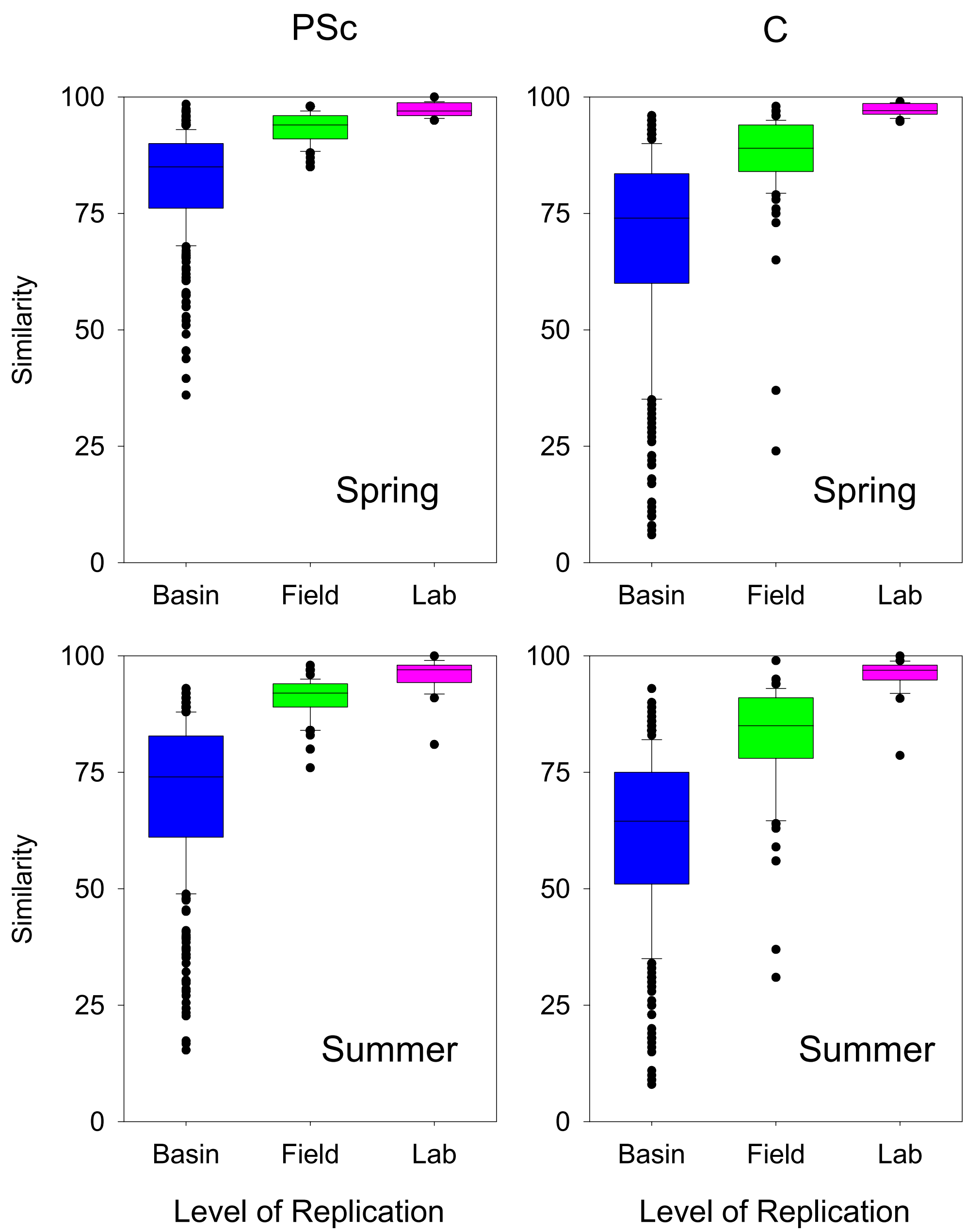

Fig. 13. Comparison of laboratory, field, and basin replicate similarity values. Data for 153-µm mesh tows, 1998 and 1999. Boxes as in Fig. 6.

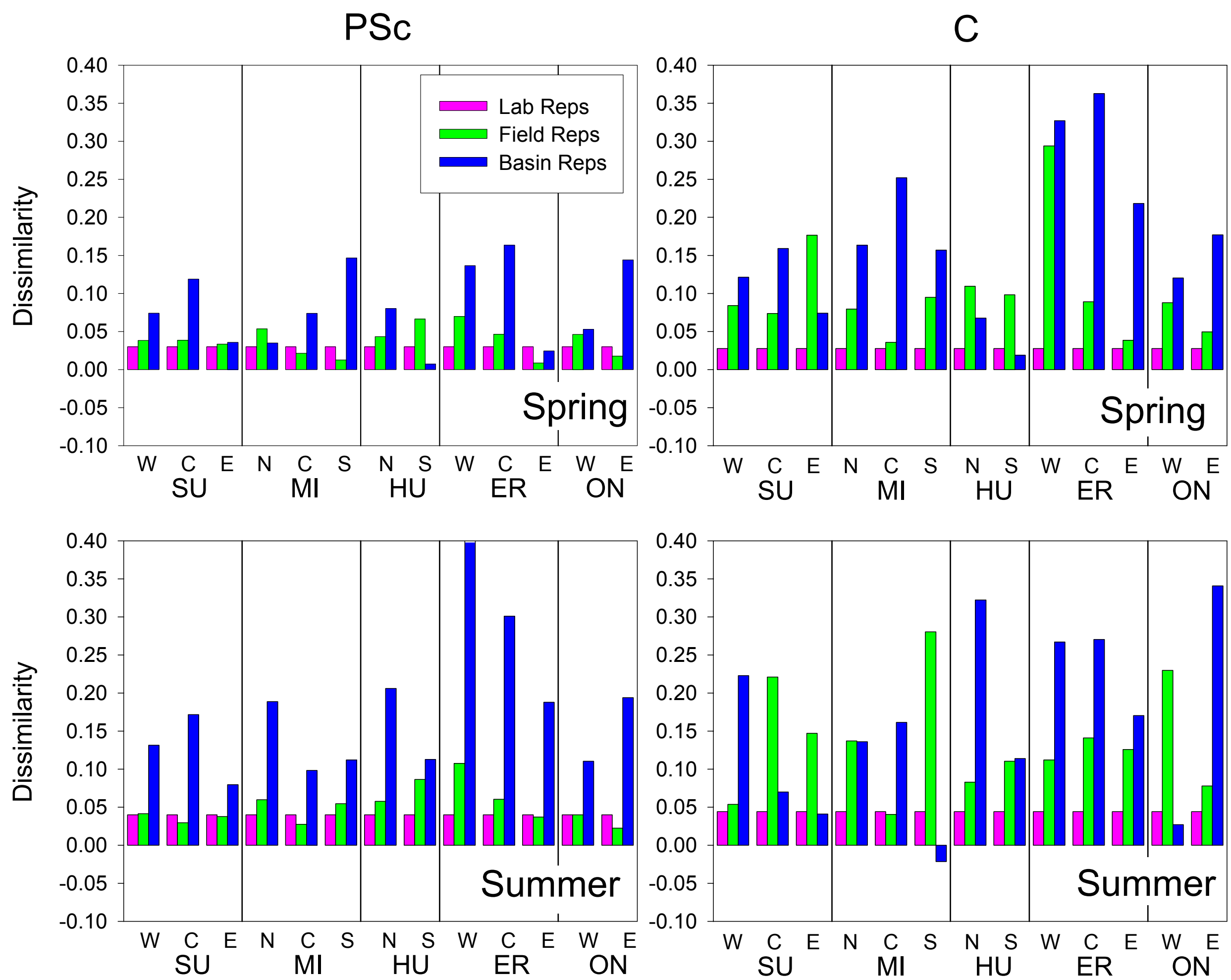

Fig. 14. Contribution to variability (as quantified by dissimilarity) of between-site heterogeneity, sample collection, and laboratory analysis.

Erie, with regard to both species composition and species densities. While such variation was high at different times in other basins, such effects were not consistently noted.

The error resulting from laboratory analyses as a percentage of overall dissimilarity (Table 17) was much higher than the error component of laboratory analyses estimated by ANOVA (Table 3). In the latter case, this source of error rarely exceeded a few percent of total within-basin variability, while dissimilarity values of duplicate laboratory analyses were approximately 10 to 25% of total within-basin dissimilarity. This in all likelihood does not represent greater variability in the taxonomical aspect of laboratory analyses (as quantified by dissimilarity values), but rather indicates that there is less variability overall involved in taxonomic analyses, as compared to estimation of densities. A direct comparison of the estimates of sources of error from ANOVA and dissimilarity analyses is not possible, however, since these two types of analysis yield quantitatively incommensurate results.

6 Discussion

6.1 DQO

Of the 184 cases examined in this study, minimum detectable differences satisfied the criterion set by the DQO in only 3 instances. While exhibiting a wide range of values, minimum detectable differences in general ranged between 40% and 190%. This means that in order for the current sampling regime to detect a change in the densities of major crustacean groups, in some cases these would have to nearly triple. Minimum detectable differences were highest for cladocerans, a group of particular ecological and management interest given their importance as fish food items. Of the regions examined, minimum detectable differences were particularly high in the western basin of Lake Erie, an area that is typically subject to high spatial heterogeneity.

While clearly not satisfying the current DQO, is the current level of sampling effort adequate to detect ecologically significant changes? Is normal interannual variability greater than the DQO criterion of 20%? Unfortunately, GLNPO does not currently possess the data necessary to address these questions. Only two years of data collected with 153-µm mesh nets is available at present, so statements about year to year variability can not be made with any confidence. While over 15 years of data collected with the 63-µm mesh net are available, as pointed out above, interannual variability in this data is confounded with variability due to diurnal vertical migration. However, a recent study (Barbiero and Tuchman, in press) was able to detect significant changes in the densities of many cladoceran species, as estimated by 63-µm mesh tows, resulting from the invasion of an exotic zooplankton predator in the mid 1980s. These changes in many cases were quite drastic, though, and it is unclear if less substantial, but still ecologically significant, changes would be detectable under the current sampling regime.

A more fundamental shortcoming of the current DQO is that it does not afford a means of assessing community-level data quality. Data quality can only be assessed individually for each of the numerous variates that collectively make up each zooplankton sample.

In spite of falling far short of the DQO, the current sampling program is apparently successful at measuring community structure, though somewhat less successful at measuring community size. Overall, relative (i.e., PSc) similarity values for within-basin comparisons were high, with most comparisons exceeding Engleberg's (1987) criterion for identical communities of 0.60. C similarity values were always lower, though this difference narrowed in the spring, compared to summer. This indicates that community structure can be assessed with some confidence, somewhat more so in the spring than in summer due to the restricted species richness during the former season.

Both the lowest similarity values, and the highest variability of similarity values, were observed in the western and central basins of Lake Erie. Because of the morphometry of these basins and the relatively high inflow (in comparison to volume) entering the western basin, these areas exhibit substantial spatial heterogeneity in many variables, so the high variability of zooplankton data is not unexpected.

6.2 Sources of variation

Both ANOVA analyses and similarity measures indicate that relatively little uncertainty is contributed by the final stages of zooplankton analysis. The amount of variance contributed to estimation of numbers of broad taxonomic categories, as estimated by ANOVA, averaged less than 5%. The amount of dissimilarity contributed by this stage of the analysis was about 3-4%, although this represented on average 20% and 12% of the total measured within basin dissimilarity for the PSc and C indices, respectively. The low variance component of this part of the analysis is not completely surprising. Duplicate counts are performed after subsamples are taken and placed in counting chambers, so discrepancies in counts would arise strictly from miscounts, rather than differences in the numbers of organisms contained in different subsamples. The counting chamber contains a circular groove which allows the sample to be enumerated essentially along a continuous transect, with most of the width of the transect remaining within the field of vision of the microscope. While discrepancies in identifications might occur between analysts, the low PSc dissimilarity values suggest that this does not happen frequently, which again is not surprising given the limited species diversity of most zooplankton samples, and the tendency for samples to be dominated by a small number of the half dozen common species.

All three measures of variance suggest that about one quarter of the total within-basin uncertainty in the zooplankton data is apparent in field replicates. Variance between field replicates contributed about 25% of within-basin variability, according to the ANOVA analysis; PSc dissimilarity values between field replicates contributed, on average, 23% of total within-basin dissimilarity, while variance between field replicates contributed 39% of total within-basin C dissimilarity. Included in this component of variance is small-scale patchiness, uncertainty associated with sample collection, and also uncertainty resulting from subsampling in the laboratory.

Station-to-station variability within a basin contributed the most variance, as quantified by all three measures, which indicates zooplankton communities vary considerably within the nominally homogeneous basins. More of this variability appears to be a result of differences in densities, rather than differences in species composition from station to station. Station-to-station variability contributed 70% of the total within-basin variability measured by ANOVA, which specifically quantifies variance in densities, while 40-60% of total within basin dissimilarity, which takes into account differences in species composition, was contributed by station-to-station differences. A comparison of PSc and C values suggests that during the spring, most of this variability was a result of differences in abundances, since C values were substantially and consistently higher than PSc values in this season. However, during summer, the relatively high PSc dissimilarity values and the lack of a substantial difference between PSc and C values indicate that species composition also varied from station to station with basins.

6.3 Controlling variation

The major source of variation in GLNPO's zooplankton data appears to be basin-scale spatial heterogeneity. The most appropriate way of reducing this source of variability would be to increase the number of stations within each basin. It is recognized that this is probably not a feasible alternative.

The error associated with field replicates contributed substantially less variability, but this stage of data generation offers more realistic possibilities for reducing overall variance. As mentioned, this variance component includes uncertainty due to subsampling in the laboratory, in addition to the uncertainty involved in sample collection and small scale spatial heterogeneity. Regression results suggest that a significant amount of uncertainty in this stage is associated with variations in flow meter readings between field replicates. This source of variability can be reduced by ensuring that all flow meters are in a good state of repair through a regular schedule of maintenance. Anomalous readings should be recognized by field personnel and should result in replacement of faulty meters. Records of meter-specific calibrations should be kept on ship so that large divergences from past calibration factors can be recognized. It is also necessary that field personnel be properly trained to ensure that meters are read correctly and that potential problems with meters are recognized early and addressed appropriately.

Other actions can be taken in the field to reduce the level of uncertainty introduced at this stage of data generation. Field personnel should ensure that zooplankton nets are thoroughly rinsed before decanting contents into sample bottles. They should also exercise care in ensuring that both net speed and depth are kept as close to those specified in the standard operating procedure as possible. The most difficult element of field sampling to control is typically the angle of the net. Interestingly, no relationship was found between variability in net angle between field replicates and levels of dissimilarity, which suggests that the impact of net angle on uncertainty might be relatively slight.

Since replicate analyses are not conducted on subsamples taken in the laboratory, the amount of variability contributed by this stage of analysis is unknown. Instead, the variability contributed by subsampling is included in estimates of between field replicate variance. Since subsampling represents a source of uncertainty that is particularly amenable to investigator control, it would be helpful to know how substantial it is. This could be accomplished by analyzing duplicate splits of a single sample. Ways of reducing uncertainty due to subsampling include ensuring that the sample is completely homogenized prior to splitting in the Folsom splitter, and ensuring that all organisms are subsequently transferred to the counting chamber.

7 References

Barbiero, R.P. 2003. Application of the Great Lakes National Program Office's Data Quality Objective to Benthos Data Generated by the Annual Water Quality Survey. US EPA, GLNPO, Chicago Il.

Barbiero, R.P. and M.L. Tuchman. 2003. Changes in the crustacean communities of Lakes Michigan, Huron and Erie following the invasion of the predatory cladoceran *Bythotrephes cederstroemi* Can. J. Fish. Aq. Sci. (*in press*).

Engelberg, K. 1987. Die Diatomeen-Zönose in eimem Mittelgegirgsbach und die Abgrenzung jahreszeitlicher Aspekte mit Hilfe der Dominanz-Identität. Arch. Hydrobiol. 110:217-236.

GLNPO 2003. Sampling and Analytical Procedures for GLNPO's Open Lake Water Quality Survey of the Great Lakes. EPA 905-R-03-002, U.S. EPA, GLNPO, Chicago Il.

Kulczynski, S. 1927. Die Pflanzenassoziation der Pieninen. Internat. Acad. Polon. Sci., Lettr. Bull., Classe Sei. Math. et Nat., ser. B. Sci. Nat. Suppl. 2:1927:57-203.

Makarewicz, J.C. 1987. Phytoplankton and zooplankton composition, abundance and distribution: Lake Erie, Lake Huron an Lake Michigan - 1983 Volume 1 and 2 U.S. Environmental Protection Agency. EPA-905/2-87-002. 183 p.

Makarewicz, J.C. 1988. Phytoplankton and zooplankton composition, abundance and distribution: Lakes Erie, Huron and Michigan - 1984 U.S. Environmental Protection Agency. EPA-905/3-88-001.

Makarewicz, J.C. and P. Bertram. 1991. Phytoplankton and zooplankton composition, abundance and distribution Lakes Erie, Huron and Michigan - 1985. U.S. Environmental Protection Agency. EPA- 905/3-85-003.

Motyka, J., B. Dobrzanski and S. Zawadzki. 1950. Preliminary studies on meadows in the southeast of the province Lublin. Univ. Mariae Curie-Sklodowska Ann. Sect. E. 5 (13):367-447.

Whittaker, R.H. 1952. A study of summer foliage insect communities in the Great Smoky Mountains. Ecol. Monogr. 22:6-31.

Whittaker, R.H. and C.W. Fairbanks 1958. A study of copepod communities in the Colombia Basins, Southeastern Washington. Ecology 39:46-63

Yan, N.D., W. Keller, N.M. Scully, D.R.S. Lean and P.J. Dillon. 1996. Increased UV-B penetration in a lake owing to drought-induced acidification. Nature. 381:141-143.